

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 13:01:31 ; Search time 39.1304 Seconds
(without alignments)
189.563 Million cell updates/sec

Title: US-09-801-784A-1
Perfect score: 173
Sequence: 1 VEKNITVTASVDPPTIDLDQDGSALPSAVALTYSPPA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	160	92.5	167	2	Q93G69	Q93G69 escherichia
2	155	89.6	170	2	Q93V20	Q93V20 escherichia
3	147	85.0	168	2	Q47123	Q47123 escherichia
4	138	79.8	170	2	Q47117	Q47117 escherichia
5	114	65.9	134	2	Q47122	Q47122 escherichia
6	113	65.3	191	16	Q9XDS3	Q9XDS3 salmonella
7	112	64.7	25	2	Q9R514	Q9R514 escherichia
8	108	62.4	166	2	Q9EXS3	Q9EXS3 burkholderi
9	108	62.4	170	2	Q9EXS2	Q9EXS2 burkholderi
10	108	62.4	184	2	Q51591	Q51591 burkholderi
11	101	58.4	142	2	Q47124	Q47124 escherichia
12	97	56.1	133	2	Q47121	Q47121 escherichia
13	97	56.1	135	2	Q47120	Q47120 escherichia
14	59.5	34.4	166	16	Q92TV3	Q92TV3 rhizobium
15	56.5	32.7	581	16	Q8X6U9	Q8X6U9 escherichia
16	55.5	32.1	2078	16	Q98K31	Q98K31 rhizobium

17	54.5	31.5	1430	16	Q9BHK1	Q9BHK1 rhizobium
18	54	31.2	330	2	Q93ST9	Q93ST9 chlorobium
19	54	31.2	433	13	Q9DE24	Q9DE24 brachydanio
20	54	31.2	516	4	Q9ULV9	Q9ULV9 homo sapien
21	54	31.2	520	4	Q9UBG7	Q9UBG7 homo sapien
22	54	31.2	552	4	Q95723	Q95723 homo sapien
23	54	31.2	1186	9	Q80211	Q80211 methanobact
24	54	31.2	2328	6	Q95M19	Q95M19 erinaceus e
25	53.5	30.9	417	16	Q9WYU7	Q9WYU7 thermococ
26	53	30.6	740	16	Q8X728	Q8X728 escherichia
27	52	30.1	118	5	Q9WYU0	Q9WYU0 dtroscophila
28	52	30.1	498	16	Q97EM4	Q97EM4 clostridium
29	52	30.1	694	16	Q97EM5	Q97EM5 clostridium
30	52	30.1	2437	6	Q95M33	Q95M33 oycoclagus
31	51.5	29.8	161	2	Q8VJ22	Q8VJ22 salmonella
32	51.5	29.8	350	17	Q8U3Y5	Q8U3Y5 pyrococcus
33	51.5	29.8	1708	16	Q8Y212	Q8Y212 anabaena sp
34	51	29.5	415	16	Q92BV8	Q92BV8 listeria in
35	51	29.5	833	12	Q65481	Q65481 bovine ente
36	51	29.5	1510	2	Q92465	Q92465 corynebacte
37	51	29.5	2358	16	Q9LIV8	Q9LIV8 streptomyce
38	50.5	29.2	194	16	Q9K809	Q9K809 bacillus ha
39	50.5	29.2	270	5	Q9NE74	Q9NE74 leishmania
40	50.5	29.2	417	16	Q98LJ7	Q98LJ7 rhizobium
41	50.5	29.2	580	16	Q8ZLF7	Q8ZLF7 salmonella
42	50.5	29.2	580	16	Q8Z242	Q8Z242 salmonella
43	50.5	29.2	1605	2	Q9X6M3	Q9X6M3 salmonella
44	50.5	29.2	3624	16	Q8Z411	Q8Z411 salmonella
45	50.5	29.2	5198	5	Q76518	Q76518 caenorhadi

ALIGNMENTS

RESULT 1
ID Q93G69 PRELIMINARY; PRT; 167 AA.
AC Q93G69;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CS4 major fimbriae subunit CSaB.
GN CSaB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=E11881A;
RA Alpboun Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and characterization of ETEC CS4 fimbriae encoding genes,
RT and their expression in Shigella flexneri 2a guinea strain CVD 1204.";
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF296132; AAK97135.1; .
SQ SEQUENCE 167 AA; 17344 MW; 94C77822A7C3A5A CRC64;

Query Match 92.5%; Score 160; DB 2; Length 167;
Best Local Similarity 88.9%; Pred. No. 2.7e-14;
Matches 32; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 VEKNITVTASVDPPTIDLDQDGSALPSAVALTYSPPA 36
DB 24 VEKNITVTASVDPPTIDLDQDGSALPSAVALTYSPPA 59
RESULT 2
ID Q93V20 PRELIMINARY; PRT; 170 AA.
AC Q93V20;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

```

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C91F-6;
RX MEDLINE=96071508; PubMed=7591145;
RT Froehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RA "Genes for CS2 pili of enterotoxigenic Escherichia coli and their
RL interchangeability with those for CS1 pili.";
RT Infect. Immun. 63:4849-4856(1995).
DR EMBL; Z47800; CAA87761.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN 24 170 COTA.
SQ SEQUENCE 170 AA; 17739 MW; DDAAL053679E3674 CRC64;

Query Match 79.8%; Score 138; DB 2; Length 170;
Best Local Similarity 79.4%; Pred.No. 3.le-11;
Matches 27; Conservative 4; Mismatches 3; Indels 0; Gaps 0

QY 2 EKNIVTASVDPTIDLLQADGSGALPSAVALTYSP 35
|||||:|||||:|||||:|||||:|||||:|
DB 25 EKNIVTASVDPTIDLLMQSDGTALPSAVNIAYLP 58

RESULT 5
Q47122 PRELIMINARY; PRT; 134 AA.
ID Q47122 PRELIMINARY; PRT; 134 AA.
AC Q47122;
QT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DD 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CsfA protein (Fragment).
GN CSFA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9B1373;
RA Gaastria W., Kusters J.G., Van Dijk L.;
RT "The major subunit gene for Csi4 fimbriae is duplicated in the Csi4
RT fimbrial operon.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X97493; CAA66124.1; -.
FT NON_TER 1
FT NON_TER 134 1
SQ SEQUENCE 134 AA; 13916 MW; 3EEFFDC5861396A4 CRC64;

Query Match 65.9%; Score 114; DB 2; Length 134;
Best Local Similarity 84.6%; Pred.No. 4.9e-08;
Matches 22; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 VDPTIDLLQADGSGALPSAVALTYSPA 36
|||||:|||||:|||||:|||||:|||||:|
DB 1 VDPTIDLLQADGSSLPATVELTYSPA 26

RESULT 6
Q9XDS3 PRELIMINARY; PRT; 191 AA.
ID Q9XDS3 PRELIMINARY; PRT; 191 AA.
AC Q9XDS3;
QT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DD 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE TSAB protein (Putative fimbrial subunit).
DE TSAB OR TCFB OR STY0346.
OS Salmonella typhi, and
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601, 28901;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=GIJ10007;
RA Hashimoto Y.;
RT "Identification of a putative fimbrial operon of Salmonella typhi.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enterica; STRAIN=RKS 3333;
RX MEDLINE=99348391; PubMed=10417651;
RA Folkesson A., Advani A., Sukupolvi S., Pfeifer J.D., Normark S.,
  Loftdal S.;
RT "Multiple insertions of fimbrial operons correlate with the evolution
  of Salmonella serovars responsible for human disease.";
RL Mol. Microbiol. 33:612-622(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
MEDLINE=21534947; PubMed=11677608;
RA Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
  Parker J., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
  Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
  Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
  Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
  Krogsh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
  Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
  Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
  enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AB029403; BAA82270.1; -;
DR EMBL; AJ242964; CAB51575.1; -;
DR EMBL; AL627266; CAD08771.1; -;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 191 AA; 20282 MW; F1DC03208131C7C0 CRC64;

Query Match 65.3%; Score 113; DB 16; Length 191;
Best Local Similarity 60.0%; Pred. No. 1e-07;
Matches 21; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 1 VEKNITVASVPTIDLLQADGSAALPSAVALTYSP 35
DB 44 VOKDITVTANIDSTIELLQADGSLPSTMKLDFWP 78

RESULT 7
AC Q9RSJ4 PRELIMINARY; PRT; 25 AA.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DB Putative colonization factor O166 (Fragment).
DS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE.
RX MEDLINE=92363580; PubMed=1354200;
RA Sommerfelt H., Grewal H.M., Svennerholm A.M., Gastra W., Flood P.R.,
  Viboud C., Bhan M.K.;
RT "Genetic relationship of putative colonization factor O166 to
  colonization factor antigen I and coli surface antigen 4 of
  enterotoxigenic Escherichia coli.";
RL Infect. Immun. 60:3799-3806(1992).
SQ SEQUENCE 25 AA; 2570 MW; 9CDFEAC96A6986A5 CRC64;

Query Match 64.7%; Score 112; DB 2; Length 25;
Best Local Similarity 92.0%; Pred. No. 1.3e-08;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEKNITVASVPTIDLLQADGSAAL 25

```

```

DB 1 VEKNITVASVPTIDLLQADGSAAL 25

RESULT 8
AC Q9EXS3 PRELIMINARY; PRT; 166 AA.
ID Q9EXS3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DB Adhesin major subunit pilin.
GN CBLA.
OC Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
  Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCPH 3025;
RX MEDLINE=21157375; PubMed=11257549;
RA Richardson J., Stead D.E., Coutts R.H.A.;
RT "Incidence of the cblA major subunit pilin gene amongst Burkholderia
  species.";
RL FEMS Microbiol. Lett. 196:61-66(2001).
DR EMBL; AJ304454; CAC18737.1; -;
SQ SEQUENCE 166 AA; 16971 MW; E05EC5C30F742FF CRC64;

Query Match 62.4%; Score 108; DB 2; Length 166;
Best Local Similarity 57.1%; Pred. No. 4.3e-07;
Matches 20; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 VEKNITVASVPTIDLLQADGSAALPSAVALTYSP 35
DB 20 VOKDITVTANVDTTLEMLADGSAALPTTMQOYLP 54

RESULT 9
AC Q9EXS2 PRELIMINARY; PRT; 170 AA.
ID Q9EXS2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DB Adhesin major subunit pilin.
GN CBLA.
OC Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
  Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BURKHOLDERIA CEPACIA COMPLEX;
RX MEDLINE=21157375; PubMed=11257549;
RA Richardson J., Stead D.E., Coutts R.H.A.;
RT "Incidence of the cblA major subunit pilin gene amongst Burkholderia
  species.";
RL FEMS Microbiol. Lett. 196:61-66(2001).
DR EMBL; AJ304455; CAC18738.1; -;
SQ SEQUENCE 170 AA; 17587 MW; 940CA79122714B16 CRC64;

Query Match 62.4%; Score 108; DB 2; Length 170;
Best Local Similarity 54.3%; Pred. No. 4.4e-07;
Matches 19; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 VEKNITVASVPTIDLLQADGSAALPSAVALTYSP 35
DB 20 VOKDITVTANIDTTEMLADGSAALPTTMQOYLP 54

RESULT 10
ID Q51591 PRELIMINARY; PRT; 184 AA.

```

```

AC Q51591;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Adhesin major subunit pilin.
GN CBIA.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC 7;
RX MEDLINE=95164502; PubMed=7532166;
RA Sujan U.S., Sun L., Goldstein R., Forstner J.F.;
RT "Cable (cbi) type II pili of cystic fibrosis-associated Burkholderia
RT (Pseudomonas) cepacia: nucleotide sequence of the cblA major subunit
RT pilin gene and novel morphology of the assembled appendage fibers.";
RL J. Bacteriol. 177:1030-1038(1995).
DR EMBL; U10244; AAA69516.1; -.
SQ SEQUENCE 184 AA; 19180 MW; FE02ECE20D2760F9 CRC64;

Query Match 62.4%; Score 108; DB 2; Length 184;
Best Local Similarity 57.1%; Pred. No. 4.8e-07;
Matches 20; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Oy 1 VEKNITVTASVDPTIDLLQADGSAALPSAVALTYSP 35
Db 20 VQKDTVTANVDITLLEMSADGSAALPTTMQGYLP 54

RESULT 11
O47124 PRELIMINARY; PRT; 142 AA.
AC Q47124;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CsuA2 protein (Fragment).
GN CSUA2.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E7476A;
RA Gastra W., Kusters J.G., Van Dijk L.;
RT "The major subunit gene for Csi4 fimbrae is duplicated in the CS14
RT fimbrial operon.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X97492; CAA66123.1; -.
DR InteP20; IPR000719; Euk_pkinase.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 142
SQ SEQUENCE 142 AA; 14498 MW; 932BF1A7989A4E7D CRC64;

Query Match 59.4%; Score 101; DB 2; Length 142;
Best Local Similarity 80.0%; Pred. No. 3.3e-06;
Matches 20; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 11 VDPITDILQADGSAALPSAVALTYSP 35
Db 1 VDPITDILQADGSAALPTAVDITLYP 25

RESULT 12
O47121 PRELIMINARY; PRT; 133 AA.
AC Q47121;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CsdA protein (Fragment).
GN CSDA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F595C;
RA Gastra W., Kusters J.G., Van Dijk L.;
RT "The major subunit gene for Csi4 fimbrae is duplicated in the CS14
RT fimbrial operon.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X97494; CAA66125.1; -.
FT NON_TER 1
FT NON_TER 133
SQ SEQUENCE 133 AA; 13865 MW; 5CF679BEC86656EE CRC64;

Query Match 56.1%; Score 97; DB 2; Length 133;
Best Local Similarity 69.2%; Pred. No. 1.1e-05;
Matches 18; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Oy 11 VDPITDILQADGSAALPSAVALTYSPA 36
Db 1 VDPKLDLQADGTSLPDSIALTYSSA 26

RESULT 13
O47120 PRELIMINARY; PRT; 135 AA.
AC Q47120;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CsdA protein (Fragment).
GN CSDA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E20719A;
RA Gastra W., Kusters J.G., Van Dijk L.;
RT "The major subunit gene for Csi4 fimbrae is duplicated in the CS14
RT fimbrial operon.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X97495; CAA66126.1; -.
FT NON_TER 1
FT NON_TER 135
SQ SEQUENCE 135 AA; 14277 MW; 88964DB09C944B6F CRC64;

Query Match 56.1%; Score 97; DB 2; Length 135;
Best Local Similarity 69.2%; Pred. No. 1.1e-05;
Matches 18; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Oy 11 VDPITDILQADGSAALPSAVALTYSPA 36
Db 1 VDPKLDLQADGTSLPDSIALTYSSA 26

RESULT 14
O92JV3 PRELIMINARY; PRT; 1869 AA.
AC O92JV3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical signal peptide protein SMC03096.
GN R02997 OR SMC03096.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```

OC Rhizobiaceae; Sinorhizobium.
 ON NCBI_TaxID=382;
 RX SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E., Dreno S., Gloux S., Godrie T., Goffeau A., Kahn D., Kias D., Lelure V., Masuy D., Pohl T., Portetelle D., Puhler A., Purnelle B., Ransperger U., Renard C., Thebaud P., Vandenberg M., Weidner S., Calbert P.;
 RT "Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
 DR EMBL: AL591792; CAC47576.1; -;
 DR InterPro: IPR000425; MIP family.
 DR PROSITE: PS00221; MIP, UNKNOWN_1.
 DR Hypothetical protein; Complete proteome.
 QY SEQUENCE 1869 AA; 19105 MW; 32851E6878708F9B CRC64;
 Query Match 34.4%; Score 59.5; DB 16; Length 1869;
 Best Local Similarity 45.2%; Pred. No. 38;
 Matches 14; Conservative 7; Mismatches 9; Indels 1; Gaps 1;
 QY 5 ITVTASVDPITDLDQAGSALPSAVALTSP 35
 DB 1323 ITVAGSAGPTLD-ISAALNALPALANTFAP 1352
 RESULT 15
 QY 08X6U9 PRELIMINARY; PRT; 581 AA.
 AC 08X6U9;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Gamma-glutamyltranspeptidase.
 GN GGT OR Z4813 OR ECS4293.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 ON NCBI_TaxID=83334;
 RX SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / ATCC 700927;
 MEDLINE=21074935; PubMed=1120651;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.U., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
 RL Nature 409:529-533 (2001).
 RX SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RIMD 0509952;
 MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T., Kohara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 DR EMBL: AE005568; AAG58553.1; -;
 DR EMBL: AP002565; BAB37716.1; -;
 DR InterPro: IPR000101; Gyltinspase.
 DR Pfam: PF01019; G_glu transpept. 1.
 DR PRINTS: PRO1210; GGTTRANSPTASE.
 DR TIGRAME: TIGR00066; g_glu trans. 1.
 DR PROSITE: PS00462; G_GLU_TRANSPEPTIDASE; 1.
 KW Complete proteome.

SQ SEQUENCE 581 AA; 62035 MW; AF7DD96C67822AA CRC64;
 Query Match 32.7%; Score 56.5; DB 16; Length 581;
 Best Local Similarity 40.0%; Pred. No. 25;
 Matches 14; Conservative 9; Mismatches 7; Indels 5; Gaps 1;
 QY 7 VTASVDPITDLDQAGSALPSAVALTSP 36
 DB 51 MVASVDATATGVDILKEGNAVDAVAAGYALA 85
 RESULT 16
 QY 098KJ1 PRELIMINARY; PRT; 2078 AA.
 AC 098KJ1;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein ml11661.
 GN ML11661.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 ON NCBI_TaxID=381;
 RX SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 MEDLINE=21062930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S., Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";
 RL DNA Res. 7:331-338 (2000).
 DR EMBL: AP002998; BAB48983.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 2078 AA; 207618 MW; BB557E48B8A5B848 CRC64;
 Query Match 32.1%; Score 55.5; DB 16; Length 2078;
 Best Local Similarity 40.6%; Pred. No. 15+02;
 Matches 13; Conservative 9; Mismatches 9; Indels 1; Gaps 1;
 QY 4 NITVTASVDPITDLDQAGSALPSAVALTSP 35
 DB 908 SVTLGSAAGCTLD-LAQPSPALPALANDFAP 938
 RESULT 17
 QY 098HK1 PRELIMINARY; PRT; 1430 AA.
 AC 098HK1;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE WD-repeat protein, beta transducin-like.
 GN ML2837.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 ON NCBI_TaxID=381;
 RX SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S., Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";

```
RL DNA RES. 7:331 338(2000).
DR EMBL; AP003009; BAB49865.1; --
DR InterPro; IPRO01680; WD40.
DR Pfam; PF00400; WD40; 12.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD00018; WD40; 3.
DR SMART; SM00320; WD40; 13.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 8.
DR PROSITE; PS00682; WD_REPEATS_2; 11.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Complete proteome.
SQ SEQUENCE 1430 AA; 153507 MW; 0B5032A1D2389E02 CRC64;

Query Match 31.5%; Score 54.5; DB 16; Length 1430;
Best Local Similarity 42.9%; Pred. No. 1.4e+02;
Matches 15; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 VERNITVTASVDPTIDILLOADGSAALPSAVALTYSP 35
Db 1045 YDSNILLTASHDGTARLDWDVDG-ALTTTLSEHYRP 1068

RESULT 18
Q93SU9
ID Q93SU9 PRELIMINARY; PRT; 330 AA.
AC Q93SU9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE B4HG.
GN B4HG.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID:1097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20433268; PubMed=10976061;
RA Xiong J., Fischer W.M., Inoue K., Nakahara M., Bauer C.E.;
RT "Molecular evidence for the early evolution of photosynthesis.";
RL Science 289:1724-1730(2000).
DR EMBL; AY004137; AAG12417.1; --
DR InterPro; IPRO01092; HUH_bas.
DR Pfam; PF01040; UbiA; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN 1.
SQ SEQUENCE 430 AA; 35477 MW; 420256115A9DB9F8 CRC64;

Query Match 31.2%; Score 54; DB 2; Length 330;
Best Local Similarity 37.8%; Pred. No. 29;
Matches 14; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 2 EKNTVTASV-----DPTIDILLOADGSAALPSAVALT 32
Db 230 EKNAALIASAVWDIAQIAIAILVAKGSTITATAVT 286

RESULT 19
Q9DE24
ID Q9DE24 PRELIMINARY; PRT; 433 AA.
AC Q9DE24;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Forkhead transcription factor c1.2.
GN FOXCl.2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID:7955;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21095190; PubMed.11165495;
RA Topczewska J.M., Topczewski J., Solnica-Krezel L., Hogan B.L.M.;
RT "Sequence and expression of zebrafish foxc1a and foxc1b, encoding
conserved forkhead/winged helix transcription factors.";
RL Mech. Dev. 100:343-347(2001).
DR EMBL; AF219950; AAG44242.1; --
DR HSSP; Q63245; 2HEH.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR PRODOM; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 433 AA; 47918 MW; 8941A9DBB9B817B9 CRC64;

Query Match 31.2%; Score 54; DB 13; Length 433;
Best Local Similarity 61.1%; Pred. No. 40;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 18 LQDGSALPSAVALTYSP 35
Db 273 LPASRAALPGSVSLTYSP 290

RESULT 20
Q9ULV9
ID Q9ULV9 PRELIMINARY; PRT; 516 AA.
AC Q9ULV9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Transcription factor RBP-L.
GN RBP-L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Saito T., Miyajima N.;
RT "Human mRNA for transcription factor RBP-L.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB027710; BAA87051.1; --
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 1.
SQ SEQUENCE 516 AA; 56594 MW; 272055B8A51C1A00 CRC64;

Query Match 31.2%; Score 54; DB 4; Length 516;
Best Local Similarity 40.7%; Pred. No. 49;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 9 ASVDPTIDILLOADGSAALPSAVALTYSP 35
Db 450 AP111PMSUVRADGLFYSAPFTYTP 476

RESULT 21
Q9UBG7
ID Q9UBG7 PRELIMINARY; PRT; 520 AA.
AC Q9UBG7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Transcription factor RBP-L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=PANCREAS;
```

RA Koyama K., Isaka S., Okamura S.;
RT "112-pa1";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Isaka S., Koyama K., Nakamura Y., Okamura S., Azuma C., Kimura T.;
RT "Human RBP-L";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB024964; BAA8832.1; -
DR EMBL; AB026048; BAA8612.1; -
DR InterPro: IPR002909; IPT_TIG.
DR Pfam: PF01833; TIG: 1
SQ SEQUENCE 520 AA; 57038 MW; BACF92459A77ACFB CRC64;

Query Match 31.2%; Score 54; DB 4; Length 520;
Best Local Similarity 40.7%; Pred. No. 49;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

9 ASVDPITDILQADGSAIPSAVALTYP 35
454 APITIPMSLVRADGLFPYSAFSTYTP 480

RESULT 22
ID 095723 PRELIMINARY; PRT; 552 AA.
AC 095723;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE DJ45C12.2 (recombining binding protein suppressor of hairless-like DE (Drosophila)).
GN RBP5UHL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bridgeman A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021578; CAAL6521.1; -
DR InterPro: IPR002909; IPT_TIG.
DR Pfam: PF01833; TIG: 1
SQ SEQUENCE 552 AA; 60422 MW; 5E2EC4B12237AF43 CRC64;

Query Match 31.2%; Score 54; DB 4; Length 552;
Best Local Similarity 40.7%; Pred. No. 53;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

9 ASVDPITDILQADGSAIPSAVALTYP 35
486 APITIPMSLVRADGLFPYSAFSTYTP 512

RESULT 23
ID 080211 PRELIMINARY; PRT; 1186 AA.
AC 080211;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Tail protein.
OS Methanobacterium phage psiM2
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxId=77048;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99009353; Pubmed=9791169;
RA Pfister P., Wasserrallen A., Stettler R., Leisinger T.;
RT "Molecular analysis of Methanobacterium phage psiM2.";
RN MO1. Microbiol. 30:233-244(1998).
RN [2]

RP SEQUENCE FROM N.A.
RA Pfister P., Wasserrallen A., Stettler R., Leisinger T.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065411; AAC27060.1; -
SQ SEQUENCE 1186 AA; 132210 MW; F26F46C4158B8F76 CRC64;

Query Match 31.2%; Score 54; DB 9; Length 1186;
Best Local Similarity 38.2%; Pred. No. 1,36+02;
Matches 13; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

1 VEXNITVTSVDPITDILQADGSAIPSAVALTYS 34
35 VEVDITETVPILETITQDSLETIPSDVEITIS 68

RESULT 24
ID 095M19 PRELIMINARY; PRT; 2328 AA.
AC 095M19;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Mannose-6-phosphate/insulin-like growth factor II receptor DE (fragment).
GN M6P/IGF2R.
OS Eriaceae; Eriaceae (Western European heathhog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Insectivora; Eriaceidae; Eriaceae; Eriaceae.
OX NCBI_TaxId=9365;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21313641; Pubmed=11420613;
RA Killian J.K., Buckley T.R., Stewart N., Munday B.L., Jittle R.L.;
RT "Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution."
RL Mamm. Genome 12:513-517(2001).
DR EMBL; AF339162; AAK71868.1; -
DR InterPro: IPR002034; AIPM/HcIt_synth.
DR InterPro: IPR000479; CIMR.
DR InterPro: IPR000562; FN_Type_II.
DR InterPro: IPR001150; Ptm_Actransf.
DR InterPro: IPR001596; Pyrophosphatase.
DR Pfam: PF00878; CIMR; 13.
DR Pfam: PF00040; fn2; 1.
DR PROSITE: PS00095; FN_Type_II; 1.
DR PROSITE: PS00815; AIFM_HOMOCTI_SYNTH_1; UNKNOWN_1.
DR PROSITE: PS00023; FIBROBLASTIN_2; UNKNOWN_1.
DR PROSITE: PS00850; GLY_RADICAL; UNKNOWN_1.
DR PROSITE: PS00387; PPA5B; UNKNOWN_1.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 2328 2328
SQ SEQUENCE 2328 AA; 257157 MW; CB64DF66CEB93911 CRC64;

Query Match 31.2%; Score 54; DB 6; Length 2328;
Best Local Similarity 47.8%; Pred. No. 2,86+02;
Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

11 VDPITDILQADGSAIPSAVALTYS 33
1533 VDQVILQVLEDGSPCKSKTGLTY 1555

RESULT 25
ID 09WYU7 PRELIMINARY; PRT; 417 AA.
AC 09WYU7;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical protein TM0476.
GN TM0476.
OS Thermotoga maritima.

Query Match 30.9%; Score 53.5; DB 16; Length 417;
Best Local Similarity 39.4%; Pred. No. 45;
Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

Search completed: January 3, 2003, 13:03:27
Job time : 41.1304 secs